

(論文審査結果の要旨)

本博士論文では、マイクロホンを意識しないで発声した音声の収録および認識アルゴリズムに関する研究について報告している。このようなマイクロホンから離れた音声を収録し認識する技術は、ハンズフリー音声認識として着目を集めており、ヒューマンインタフェースの画期的な改善につながるとして期待されている。このハンズフリー音声認識をマイクロフォンアレーを用いて実現している。学術論文として4件(第1著者3件、第2著者1件)発表しており、また、国際会議では8件(第1著者4件、第2著者4件)の発表を行っている。また、理論だけでなく、実環境での評価や、実際に実時間で動作するデモシステムも構築している点も高く評価できる。以下、論文のポイントについて述べる。

まず、複数の音源の位置推定および発話者の位置推定を行うために、CSP(Cross-power Spectrum Phase analysis)法を拡張したCSP加算法を提案し、実環境でも複数音源位置を良好に推定できることを示している。残響などのある実環境でも安定して動作するCSP加算法を提案したことは高く評価できる。

次に、遠隔発話音声を高品質に受音するには、CSP加算法で話者位置を同定した後に、発話者にマイクロホンアレーの指向特性を向けることが必要となる。この論文では、マルチビームフォーミング(J.L.Flanaganらによって1993年に提案)を用いて研究を進展させ、実環境における諸問題を解決している。直接音のみからの音源推定法や従来のマルチビームフォーミング法よりも高品質な受音のアルゴリズムが確立されている。これを、実環境におけるSNRおよび音声認識実験によって確かめている。とくにマイクロホンアレーを用いたハンズフリー音声認識として、現時点では、最も頑健なアルゴリズムであると思われる。

最後に、上記の研究成果を用いて、遠隔発話音声を高品質に受音し、さらに話者の映像情報を取得するために、マイクロホンアレーとビデオカメラを自動的に制御するシステムを構築して、実時間で動作することを確認し、さらに音声認識と結合して、音声翻訳システムの音声認識部を構築している。

以上述べたように、マイクロフォンアレーを用いたハンズフリー音声認識の問題を、理論から実環境での評価、さらに、デモシステムの構築も行っており、研究としてよくまとまっている。この分野は、音声インタフェースの新しい有望な分野であり、この先導的な研究は、情報科学、とくに音情報処理の分野に多大な貢献をしたと評価できる。平成13年7月2日開催した公聴会の結果も参考にして、本博士論文の審査を行い、本論文は博士(工学)の学位論文として十分な価値があるものと判断した。